Received: 1 October 2024 | Accepted: 13 October 2024

DOI: 10.1002/mpr.70006

ORIGINAL ARTICLE

WILEY

# A prediction model for differential resilience to the effects of combat-related stressors in US army soldiers

Ronald C. Kessler[1] | Robert M. Bossarte[2,3] | Irving Hwang[1] | Alex Luedtke[4,5] | James A. Naifeh[6] | Matthew K. Nock[7] | Maria Petukhova[1] | Ekaterina Sadikova[1,8] | Nancy A. Sampson[1] | Erik Sverdrup[9] | Jose R. Zubizarreta[1] | Stefan Wager[10] | James Wagner[11] | Murray B. Stein[12] | Robert J. Ursano[6]

[1]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA

[2]Department of Psychiatry and Behavioral Neuroscience, Morsani School of Medicine Tampa, University of South Florida, Tampa, Florida, USA

[3]Center for Mental Health Outcomes Research, Central Arkansas VA Medical Center, North Little Rock, Arkansas, USA

[4]Department of Statistics, University of Washington, Seattle, Washington, USA

[5]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

[6]Department of Psychiatry, Center for the Study of Traumatic Stress, Uniformed Services University School of Medicine, Bethesda, Maryland, USA

[7]Department of Psychology, Harvard University, Cambridge, Massachusetts, USA

[8]Department of Social and Behavioral Sciences, Harvard Chan School of Public Health, Boston, Massachusetts, USA

[9]Department of Econometrics & Business Statistics, Monash University, Melbourne, Victoria, Australia

[10]Graduate School of Business, Stanford University, Stanford, California, USA

[11]Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor, Ann Arbor, Michigan, USA

[12]Departments of Psychiatry and Family Medicine and Public Health, University of California San Diego, La Jolla, California, USA

**Correspondence**

Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA, 02115, USA.
Email: kessler@hcp.med.harvard.edu

**Funding information**

National Institute of Mental Health, Grant/Award Number: U01MH087981; U.S. Department of Defense, Grant/Award Number: HU0001-15-2-0004; U.S. Army, Grant/Award Number: U01MH087981

## Abstract

**Objectives:** To develop a composite score for differential resilience to effects of combat-related stressors (CRS) on persistent DSM-IV post-traumatic stress disorder (PTSD) among US Army combat arms soldiers using survey data collected before deployment.

**Methods:** A sample of $n = 2542$ US Army combat arms soldiers completed a survey shortly before deployment to Afghanistan and then again two to three and 8–9 months after redeployment. Retrospective self-reports were obtained about CRS. Precision treatment methods were used to determine whether differential resilience to persistent PTSD in the follow-up surveys could be developed from pre-deployment survey data in a 60% training sample and validated in a 40% test sample.

**Results:** 40.8% of respondents experienced high CRS and 5.4% developed persistent PTSD. Significant test sample heterogeneity was found in resilience ($t = 2.1$, $p = 0.032$), with average treatment effect (ATE) of high CRS in the 20% least

resilient soldiers of 17.1% (SE = 5.5%) compared to ATE = 3.8% (SE = 1.2%) in the remaining 80%. The most important predictors involved recent and lifetime pre-deployment distress disorders.

**Conclusions:** A reliable pre-deployment resilience score can be constructed to predict variation in the effects of high CRS on persistent PTSD among combat arms soldiers. Such a score could be used to target preventive interventions to reduce PTSD or other resilience-related outcomes.

**KEYWORDS**
military deployment, post-traumatic stress disorder (PTSD), precision treatment, resilience

## 1 | INTRODUCTION

Post-traumatic stress disorder (PTSD) has been described as the signature mental disorder of war (Paulson & Krippner, 2007). The syndrome was known as early as at the time of Homer's *Iliad* and has been called by many names since then, including *Soldier's Heart* in the US Civil War, *Shell Shock* in World War I, and *Battle Fatigue* in World War II (Jones, 2006). The contemporary diagnosis of PTSD was first codified in the third edition of the American Psychiatric Association Diagnostic and Statistical Manual (DSM) of Mental Disorders in 1980 and has gone through a series of refinements since that time both in the DSM system (Bovin et al., 2016) and in the World Health Organization's International Classification of Diseases (Maercker et al., 2022). Although many other types of traumatic experience exist than those associated with war (Benjet et al., 2016), and only a minority of cases of PTSD occur among military veterans (Kessler et al., 2017), military personnel and veterans have consistently been found to have a significantly higher prevalence of PTSD than the general population (Sabé et al., 2024).

Numerous attempts have been made to develop resilience programs to prevent the onset of PTSD and other stress-related disorders in the general population. However, experimental evaluations have for the most part shown that these programs are ineffective (Leppin et al., 2014). More promising results have been found for resilience training programs administered pre-deployment to military and emergency services workers (Doody et al., 2021). But successful programs of this sort are labor-intensive. The feasibility of implementing these programs in military personnel would be increased by developing a method to predict prior to deployment which military personnel are at highest risk of combat-related PTSD.

Observational studies have been carried out that provide some information about the pre-deployment predictors of combat-related PTSD that could be used to develop composite prediction models (Xue et al., 2015). And some research has been carried out showing that subsequent PTSD can be predicted with relatively good accuracy from models based on these pre-deployment variables (Papini et al., 2023; Schultebraucks et al., 2021; Wild et al., 2016). The most ambitious study of this sort was carried out by Papini et al. (2023) in a large ($n$ = 4771) pre-post deployment sample of soldiers in three US Army combat brigade teams that participated in the Army STARRS study (Ursano et al., 2014), a large-scale epidemiological-neurobiological study of risk and protective factors for Army suicides and related outcomes. This sample was administered a self-report pre-deployment survey to obtain information about risk factors shortly before deploying to Afghanistan. Follow-up surveys were then administered after the soldiers returned from deployment to determine if they met criteria for PTSD. This data array allowed a parsimonious machine learning model to be developed from pre-deployment variables that predicted post-deployment PTSD with good accuracy (test sample AU-ROC = 0.74). The one-third of soldiers with highest predicted risk based on their profile of pre-deployment characteristics accounted for 62.4% of all post-deployment PTSD.

One limitation of models like this one is that they predict PTSD rather than resilience to combat-related stressors (CRS). The distinction is important because PTSD is a joint function of trauma exposure and emotional reactivity once trauma exposure occurs. It is unclear which of these components (i.e. trauma exposure or emotional reactivity) or their combination was predicted by the Papini et al. model. If our goal in targeting soldiers for a resilience training intervention is to focus on the soldiers most likely to develop PTSD if exposed to CRS, it would muddy the water to include predictors of PTSD along with predictors of resilience, as we know that the predictors of the two can be quite different. For example, soldiers in combat arms occupations (e.g., infantry, field artillery) are known to be more highly exposed to CRS than are those in combat support (e.g., combat engineers, maintenance) or service support (e.g., communications, supply support) occupations, resulting in highest PTSD prevalence occurring among soldiers with combat arms occupations (Hines et al., 2014). However, differential resilience is quite a different matter, as shown by the fact that risk of PTSD after exposure to CRS is significantly lower among soldiers with combat arms occupations than either combat support and service support occupations (Kok et al., 2020).

While this problem of confounding could be addressed in part by focusing on soldiers in combat arms occupations to determine which of them is least resilient, a remaining challenge in developing a model for resilience is that resilience, unlike PTSD, cannot be observed directly but instead needs to be inferred from aggregate associations of the sort described in the last paragraph. For example, our assertion

that soldiers in combat support or service support occupations are, on average, less resilient than soldiers in combat arms occupations was based on evidence that the predictive association of combat stressor exposure with post-deployment PTSD is stronger among the former than the latter soldiers. More generally, the existence of differential resilience with respect to a given pre-deployment measure is inferred in this same way from the existence of a statistically significant interaction between that measure and subsequent exposure to CRS in predicting post-deployment PTSD.

The logic here is similar to that of estimating heterogeneous treatment effects (HTE) in observational studies of comparative treatment effects when we have a large number of potential modifiers (Varadhan & Seeger, 2013). Two main issues arise in such exercises. First, unlike in the situation with randomized treatment experiments in which we can reasonably assume that treatment assignment (or, in our case, assignment to exposure to CRS) is random with respect to baseline predictors of HTE, this assumption cannot be made in observational studies. This problem is addressed with a variety of analytic methods of matching or weighting that adjust for the possibility of nonrandom assignment to stressor exposure with respect to measured baseline predictors (Chattopadhyay et al., 2020).

Second, even when treatment assignment can be assumed to be random with respect to baseline predictors, a question arises how best to search for an optimal multivariable prediction rule when the outcome of differential resilience is unobserved. One simple approach is to estimate a model that includes interactions of many baseline predictors with the dummy variable(s) for treatment assignment in a multiple regression analysis. When multiple interactions are statistically significant and there are only two treatment arms, predicted outcome scores for each participant in the study can be estimated twice using counter-factual logic: once assuming the participant was assigned to the control group (i.e., exposed to CRS); and the other time assuming that the participant was assigned to the intervention group (i.e., not exposed to CRS). These two individual-level treatment-specific predicted outcome scores can then be compared to estimate individual-level resilience. And these within-individual difference scores can then be compared across individuals to estimate differential resilience (VanderWeele et al., 2019).

This counter-factual approach can lead to over-fitting when many interactions are considered, but this problem can be minimized by the use of independent training, tuning, and testing samples (Abadie et al., 2018). Even when this is done, though, another problem is that accuracy requires correct specification of both the (possibly nonlinear) main effects and the (possibly complex nonlinear and higher order) interactions. This problem can be addressed, by estimating interactions directly either by using the individual-level difference scores as an outcome (Murphy, 2003; Robins, 2004) or by using machine learning methods that achieve the same end of avoiding the need to specify main effects correctly and allow the data-driven estimation of nonlinear and higher-order interactions (Kennedy, 2023; Luedtke & van der Laan, 2017; Wager & Athey, 2018).

We explore the use of a machine learning approach of the latter sort to develop a prediction model that generates a composite score for resilience to CRS in an extension of the Papini et al. analysis of the STARRS data. In doing this, we make use of the fact that the post-deployment follow-up surveys administered to STARRS respondents asked about exposure to CRS, allowing us to examine interactions between determinants of differential resilience assessed in the pre-deployment survey and subsequent exposure to these stressors in predicting post-deployment PTSD. The analysis approach we use to estimate differential resilience is one that extends the widely-used Random Forests (RF) machine learning method (Breiman, 2001) both to balance for nonrandom associations of pre-deployment predictors with CRS and to generate estimates that combine information about complex interactions across a wide range of predictors to estimate the composite resilience score (Wager & Athey, 2018).

## 2 | METHODS

### 2.1 | Participants

As noted above, data come from the Army STARRS (Study to Assess Risk & Resilience in Servicemembers) study (Ursano et al., 2014), a large-scale epidemiological-neurobiological study of risk and resilience factors for Army suicides and related outcomes. STARRS includes several sub-studies, one of them being the Pre-Post Deployment Survey (PPDS), a four-wave self-report panel survey of soldiers from three Brigade Combat Teams (BCTs) that was used by Papini et al. (2023) to estimate their model of post-deployment PTSD. PPDS respondents were administered a comprehensive survey of hypothesized risk and resilience factors 1–2 months before deployment (T0; October 2011-February 2012), a brief follow-up survey 2–3 weeks after returning from deployment (T1; September 2012-February 2013) that asked about exposure to a checklist of CRS, and then two subsequent surveys 2–3 months (T2; October 2012-March 2013) and 8–9 months (T3; June 2013-May 2014) after returning from deployment that assessed PTSD.

PPDS participants were on duty the day of T0 and provided signed informed consent to participate in the longitudinal study. Recruitment, consent, and data collection procedures detailed elsewhere (Kessler, Colpe, et al., 2013) were approved by the Human Subjects Committees of the Uniformed Services University of the Health Sciences (Bethesda, Maryland) for the Henry M. Jackson Foundation (the primary grantee), the Institute for Social Research at the University of Michigan (Ann Arbor, Michigan, the organization collecting the survey data), Harvard Medical School (Boston, Massachusetts), and the University of California San Diego (La Jolla, California).

Our sample was somewhat different from the sample used by Papini et al. in several ways. First, we focused on enlisted Regular Army soldiers, excluding officers because of their low PTSD rates. Second, we focused on PPDS respondents with combat arms occupations (52.6% of all PPDS respondents) given that, as noted in the introduction, combat arms soldiers are known to have much higher rates of exposure to CRS and associated PTSD than soldiers in combat support and service support occupations. Third, we limited

the analysis to PPDS respondents who completed both the T2 and T3 follow-up surveys because we wanted to focus on *persistent* PTSD (i.e., PTSD found both at T2 and at T3).

The full baseline PPDS sample consisted of 9488 T0 soldiers (95.3% of unit members on duty the day of T0), 86.0% of whom ($n$ = 8558) both completed the T0 survey and consented to link survey responses to administrative records. 90.5% of the latter ($n$ = 7742) subsequently deployed to Afghanistan and 71.4% of those that deployed ($n$ = 5528) completed both T2 and T3 surveys, including $n$ = 2590 enlisted soldiers with a Combat Arms duty assignment. We excluded a small number of these soldiers because they either withdrew consent ($n$ = 27) or had additional deployments as of T3 ($n$ = 21), leaving $n$ = 2542 respondents in the analysis. Weights adjusted for baseline differences between the sample and the population on administrative variables. Additional details about PPDS design, sampling, and weighting are reported elsewhere (Kessler, Colpe, et al., 2013; Kessler, Heeringa, et al., 2013).

## 2.2 | Measures

### 2.2.1 | Post-traumatic stress disorder

T0 respondents completed a computerized 9-question screening version of the PTSD Checklist (PCL; Weathers et al., 1993) to assess 30-day prevalence of PTSD. This screen was calibrated in an earlier Army STARRS survey sample (Kessler, Santiago, et al., 2013) to generate DSM-IV diagnoses based on blinded clinical reappraisal interviews with the Structured Clinical Interview for DSM-IV (SCID; Kessler, Santiago, et al., 2013). The same screening scale and calibrated diagnostic threshold were used to define PTSD at T2 and T3. This threshold was found in the Army STARRS calibration study to have moderate concordance (AU-ROC = 0.75) with blinded SCID diagnoses (Kessler, Santiago, et al., 2013). Our outcome of *persistent* PTSD was defined as scoring above the calibrated diagnostic threshold at both T2 and T3.

### 2.2.2 | Combat-related stressors

PPDS respondents were asked at T1 about 10 CRS selected from the Deployment Risk & Resilience Inventory-2 (DRRI-2; Vogt et al., 2013) and the Walter Reed Army Institute of Research Land Combat Study (Hoge et al., 2004). These CRS included experiences both in combat (e.g., wounded by the enemy) and post-combat (e.g., exposed to the sights, sounds, or smells of severely wounded or dying people), each assessed using a 5-category frequency of occurrence response scale (0, 1, 2–4, 5–9, 10+ times) that was coded to midpoint values and a top code of 10 (i.e., 0, 1, 3, 7, 10) for purposes of calculating means. As detailed below, we carried out a preliminary analysis of the associations among these reports and their joint associations with persistent PTSD to define a dichotomous measure of high exposure to CRS.

### 2.2.3 | Pre-deployment predictors of resilience

The T0 survey assessed 410 hypothesized predictors of risk and resilience based on prior research (Table S1). Key among these were measures of lifetime and recent DSM-IV mental disorders assessed with the Composite International Diagnostic Interview Screening Scales (CIDI-SC; U.S. Department of Veterans Affairs, 2023). As detailed elsewhere (Kessler, Santiago, et al., 2013), an Army STARRS clinical reappraisal study found moderate to good concordance between diagnoses based on the CIDI-SC and diagnoses based on blinded clinical interviews with the SCID (AU-ROC = 0.70–0.79). Other hypothesized T0 predictors included lifetime and recent self-injurious thoughts and behaviors, childhood family adversities, exposure to other lifetime traumatic experiences, lifetime and recent concussions and traumatic brain injuries, 12-month acute and chronic stressors, 12-month treatments for neuropsychiatric or psychological problems, 30-day neurophysiological or psychological symptoms, personality, social networks and supports, Army career characteristics, and socio-demographics.

As noted in the introduction, predicted resilience is expected to differ from predicted PTSD in that a machine learning model for predicted PTSD combines information about predictors of differential likelihood of being exposed to traumatic stress with information about predictors of differential resilience once exposed. To make this distinction clearer in evaluating predictor effects, we generated a 10-fold cross-validated predicted persistent PTSD score in our training sample using all the above pre-deployment variables as potential predictors. We then added this composite variable to the predictor set when training a model to estimate differential resilience to the effects of high CRS on persistent PTSD. The existence of this composite predictor variable allowed us to determine in the post-analysis investigation of variable importance the extent to which differential probability of PTSD was implicated in defining differential resilience.

## 2.3 | Statistical analysis

### 2.3.1 | Defining high CRS

We carried out an exploratory factor analysis of the 10 CRS measures and then inspected the gross association of a summary factor-based CRS score with persistent PTSD. Based on this inspection, a dichotomous measure was created of high CRS.

### 2.3.2 | Estimating average effects of CRS on persistent PTSD

As noted in the introduction, although intervention effects cannot be estimated unequivocally with observational data, our analyses implicitly assumed as a first approximation that any nonrandom differences in risk of exposure to high CRS can be corrected by controlling for the baseline covariates described above (Imbens &

Rubin, 2015; Rosenbaum & Rubin, 1983). A propensity score approach (Desai & Franklin, 2019) was used to adjust for significant differences of this sort in baseline predictors between soldiers who were and were not exposed to high CRS. Each soldier *i* was assigned a weight of $1/p_i$, where the propensity score $p_i$ was the estimated probability of exposure to high CRS for a soldier with that multivariable profile of pre-deployment characteristics generated using the RF ML method (Breiman, 2001).

In addition, a predicted probability of persistent PTSD after returning from deployment was estimated for each soldier given baseline covariates separately assuming that the soldier was and then again assuming that he was not exposed to high CRS using a separate RF analysis in each case. The estimated average treatment effect (ATE) of exposure to high CRS on persistent PTSD was then defined as the difference in predicted probability of persistent PTSD in the presence versus absence of high CRS after adjusting for nonrandom exposure to high CRS with respect to the range of variables assessed in the baseline survey. ATE was estimated using a doubly robust method (Funk et al., 2011; Hernán & Robins, In Press) that combined information from the two RF analyses. All steps were performed using Generalized Random Forests (GRF), an ML approach that expands on RF (Athey et al., 2019; Wager & Athey, 2018) with a focus on estimating effects on pre-deployment predictors on resilience adjusting for measured confounders. These analyses used the *grf R* package (Athey et al., 2024). A didactic overview of the GRF method is presented in a forthcoming special review paper in this journal (Sverdrup et al., In Press).

### 2.3.3 | Estimating differential resilience to the effects of high CRS

Differential resilience is defined as variation in the estimated conditional average treatment effect (CATE); that is, the expected effect of exposure to high CRS on persistent PTSD for an individual soldier conditional on that soldier having a specific combination of baseline covariate values. CATEs were first estimated using *grf* in a 60% training sample and then evaluated in a remaining 40% test sample. The evaluation entailed imputing CATEs separately to each soldier in the test sample based on the model coefficients estimated in the training sample, dividing the test sample into low and high predicted resilience subgroups defined by these estimated CATEs, and then estimating ATEs within each test sample subgroup using data from the test sample to determine whether the ATE was, in fact, higher in the subgroup predicted to be less resilient than in the subgroup predicted to be more resilient. But where should the threshold be for dividing the predicted CATE into these different subgroups? This question is resolved in *grf* by using a sequential method that estimates ATE separately for the *q-th* fraction of soldiers in the test sample with the lowest predicted CATE based on the training-sample model and then increases the value of *q* in sequential calculations to generate a Targeting Operator Characteristic (TOC) curve that compares these subgroup ATE estimates with the ATE estimated for the total test sample over all values of *q*. Area under the TOC curve can then be calculated and evaluated to determine the statistical significance of heterogeneity in resilience based on the model developed in the training sample (Yadlowsky et al., 2024).

### 2.3.4 | Predictor importance

Predictor importance in defining CATEs was examined using the tree Shapley Additive Explanations (SHAP) method (Lundberg & Lee, 2017) implemented in the *xgboost R* package (Chen et al., 2024). Individual-level values of predicted CATE based on *grf* were the outcome in this analysis. The predictors were the T0 variables used to define predicted CATE. The SHAP value analysis estimated the implications for predicting the outcome of changing each predictor from its observed score to the sample mean averaged across all logically possible permutations of other predictors in the XGBoost model (Chen & Guestrin, 2016). A higher mean absolute SHAP value suggests a more important predictor. Proportional mean absolute SHAP values ($SHAP_P$) were calculated by dividing mean absolute SHAP values of important predictors and classes of predictors by the mean absolute SHAP value of the entire model. Bee swarm plots were used to identify dominant directions of associations.

All analyses exploring the possibility of heterogeneity in resilience were carried out using *R* version 3.6.3 (R Core Team, 2018). Data management was implemented with SAS Version 9.4 (SAS Software 9.4 edn., 2013). Analyses were carried out January-June 2024.

## 3 | RESULTS

### 3.1 | Sample distribution

All combat arms soldiers at the time of study were male. Their median age was 24 (interquartile range: 21–28; Table 1). 72.7% were Non-Hispanic White, 7.7% Non-Hispanic Black, 12.2% Hispanic, and 7.5% other (including nonresponse) race-ethnicity. The majority (54.7%) were married or never married (36.3%), with the remaining 9.0% previously married. The great majority (81.4%) had a high school degree, with smaller proportions having less than a high school degree (13.1%, including GED and alternative certification) or more than high school education (5.6%, including some college or college graduate). 36.9% had a junior enlisted rank (private through private first class), while 17.5% were noncommissioned officers (corporal through staff sergeant) and the remaining 8.5% were senior noncommissioned officers (sergeant first class or master sergeant). 36.9% were in service 24 months or less at the time of deployment compared to 17.5% in service 2–4 years, 24.6% 4–8 years, and the remaining 21.9% more than 8 years. Most respondents had either 0 (42.7%) or 1 (27.2%) prior combat deployments.

**TABLE 1** Socio-demographic and Army career characteristics of all respondents in the STARRS PPDS survey who were enlisted Regular Army soldiers that completed both the T2 and T2 follow-up assessments by presence or absence of persistent PTSD at T2 and T3.

| | | | With or without persistent PTSD at T2 and T3 | | | |
|---|---|---|---|---|---|---|
| | Total | | With | | Without | |
| | (*n* = 2542) | | (*n* = 132) | | (*n* = 2410) | |
| | % | (SE) | % | (SE) | % | (SE) |
| Age | | | | | | |
| 18–21 | 27.8 | (0.9) | 21.3 | (3.5) | 28.2 | (0.9) |
| 22–24 | 30.0 | (0.9) | 30.6 | (4.2) | 30.0 | (1.0) |
| 25–27 | 21.0 | (0.9) | 24.2 | (4.0) | 20.8 | (0.9) |
| 28+ | 21.1 | (0.9) | 23.8 | (4.0) | 21.0 | (0.9) |
| $\chi^2_3$ (*p*) | | | | 3.7 (0.30) | | |
| Gender[b] | | | | | | |
| Male | 100.0 | (–) | 100.0 | (–) | 100.0 | (–) |
| Race/ethnicity[c] | | | | | | |
| Non-Hispanic Black | 7.7 | (0.6) | 6.4 | (2.5) | 7.8 | (0.6) |
| Non-Hispanic White | 72.7 | (0.9) | 71.8 | (4.2) | 72.7 | (1.0) |
| Hispanic | 12.2 | (0.7) | 15.8 | (3.4) | 12.0 | (0.7) |
| Other | 7.5 | (0.5) | 6.0 | (2.1) | 7.6 | (0.6) |
| $\chi^2_3$ (*p*) | | | | 1.8 (0.61) | | |
| Education | | | | | | |
| GED or equivalent | 13.1 | (0.7) | 23.1 | (3.9) | 12.5 | (0.7) |
| High school diploma | 81.4 | (0.8) | 72.2 | (4.2) | 81.9 | (0.8) |
| More than high school | 5.6 | (0.5) | 4.7 | (2.2) | 5.6 | (0.5) |
| $\chi^2_3$ (*p*) | | | | 6.7[a] (0.035) | | |
| Marital status | | | | | | |
| Married | 54.7 | (1.0) | 62.5 | (4.3) | 54.3 | (1.1) |
| Previously married | 9.0 | (0.7) | 10.6 | (2.9) | 8.9 | (0.7) |
| Never married | 36.3 | (1.0) | 26.9 | (3.8) | 36.8 | (1.0) |
| $\chi^2_3$ (*p*) | | | | 5.8 (0.06) | | |
| Years in service | | | | | | |
| 0–2 | 36.9 | (1.0) | 22.4 | (3.4) | 37.8 | (1.0) |
| 2–4 | 17.5 | (0.8) | 24.3 | (3.8) | 17.1 | (0.8) |
| 5–8 | 24.5 | (1.0) | 36.5 | (4.5) | 23.8 | (1.0) |
| 9+ | 21.1 | (0.9) | 16.8 | (3.7) | 21.3 | (0.9) |
| $\chi^2_3$ (*p*) | | | | 18.9[a] (<0.001) | | |
| Rank | | | | | | |
| Junior enlisted | 34.4 | (0.9) | 20.4 | (3.3) | 35.2 | (1.0) |
| Noncommissioned officer | 57.1 | (1.0) | 74.5 | (3.8) | 56.1 | (1.0) |
| Senior noncommissioned officer | 8.5 | (0.6) | 5.1 | (2.3) | 8.7 | (0.6) |
| $\chi^2_3$ (*p*) | | | | 18.2[a] (<0.001) | | |

TABLE 1 (Continued)

| | Total | | With or without persistent PTSD at T2 and T3 | | | |
|---|---|---|---|---|---|---|
| | | | With | | Without | |
| | (*n* = 2542) | | (*n* = 132) | | (*n* = 2410) | |
| | % | (SE) | % | (SE) | % | (SE) |
| Number of prior combat deployments | | | | | | |
| 0 | 42.7 | (1.0) | 28.4 | (3.8) | 43.5 | (1.0) |
| 1 | 27.2 | (0.9) | 37.4 | (4.5) | 26.6 | (1.0) |
| 2+ | 30.1 | (1.0) | 34.2 | (4.5) | 29.9 | (1.0) |
| $\chi^2_3$ (*p*) | | | | 12.8[a] (0.002) | | |

Abbreviations: GED, General Educational Development Test; PPDS, Pre-Post Deployment Study; SE, standard error; T2, time 2; T3, time 3.

[a]The 0.05 level, two-sided test.

[b]Gender, based on Army administrative records, noting that this might be different from sex at birth.

[c]Race/ethnicity, based on Army administrative records, which, in turn, are based on soldier self-reports within the categories specified.

TABLE 2 Exploratory factor analysis of 10 combat-related stressors among enlisted PPDS soldiers with Combat Arms Military Occupational Specialties that completed both the T2 and T2 follow-up assessments (*n* = 2542).

| | Any exposure | | Mean number of exposures…[a] | | | | |
|---|---|---|---|---|---|---|---|
| | % | (SE) | In total sample mean | (SE) | Among exposed mean | (SE) | Unrotated factor loadings |
| Dangerous duty (e.g., combat patrols, route clearance, clearing buildings, disarming civilians, working in areas that had IEDs) | 91.5 | (2.1) | 8.5 | (0.2) | 9.3 | (0.1) | 0.48 |
| Fired rounds at enemy or took enemy fire (either direct or indirect) | 85.1 | (1.4) | 6.2 | (0.2) | 7.3 | (0.2) | 0.64 |
| Had "close call" (e.g., equipment shot off body, IED exploded near you) | 52.7 | (1.6) | 1.7 | (0.1) | 3.2 | (0.1) | 0.72 |
| You were injured | 28.1 | (1.2) | 0.3[b] | (0.0) | 1.0[b] | (0.0) | 0.50 |
| Had unit members seriously wounded or killed | 78.3 | (2.4) | 2.8 | (0.1) | 3.5 | (0.1) | 0.54 |
| Direct responsibility for the death of an enemy combatant | 32.2 | (1.5) | 1.2 | (0.1) | 3.9 | (0.2) | 0.54 |
| Saved life of another soldier or civilian | 18.6 | (1.0) | 0.5 | (0.0) | 2.6 | (0.2) | 0.51 |
| Saw homes or villages destroyed or people begging for food | 64.4 | (1.8) | 5.1 | (0.2) | 7.9 | (0.1) | 0.52 |
| Exposed to sights/sounds/smells of wounded/dying people, population violence, or non-combatant maltreatment | 75.7 | (2.0) | 3.8 | (0.1) | 5.1 | (0.1) | 0.74 |
| Any other highly stressful experience | 70.8 | (1.1) | 4.4 | (0.1) | 6.1 | (0.1) | 0.65 |

Abbreviations: IED, improvised explosive device; PPDS, Pre-Post Deployment Study; PTSD, post-traumatic stress disorder; SE, standard error of estimate; T0, time 0.

[a]Injury was coded as a 0–1 dichotomy, while the other combat-related stressors were coded in the range 0–10, where the original values with their recoded values were 0 (recoded 0), 1 (recoded 1), 2–4 (recoded 3), 5–9 (recoded 7), and 10+ (recoded 10).

[b]The mean among the exposed is 1, by definition, for injury, as the survey did not ask respondents who reported injuries the number of times they were injured.

Estimated prevalence of persistent post-deployment PTSD was 5.4% (SE = 0.5%). The distribution of socio-demographics did not differ significantly in subsamples of respondents with and without persistent PTSD. However, persistent PTSD was inversely related to time in service ($\chi^2_3$ = 18.9, $p < 0.001$) and number of prior combat deployments ($\chi^2_2$ = 12.8, $p$ = 0.002) and was higher among non-commissioned officers than either junior enlisted soldiers or senior noncommissioned officers ($\chi^2_2$ = 18.2, $p < 0.001$).

## 3.2 | Defining high CRS

Prevalence of the 10 CRS ranged from a high of 91.5% for having dangerous duty assignments (Mean = 8.5 on the 0–10 scale in the total sample and 9.3 among those exposed to this stressor at least once) to 18.6% for saving the life of another soldier or noncombatant (Mean = 0.5 in the total sample and 2.6 among those exposed at least once; Table 2). Exploratory factor analysis using the parallel analysis

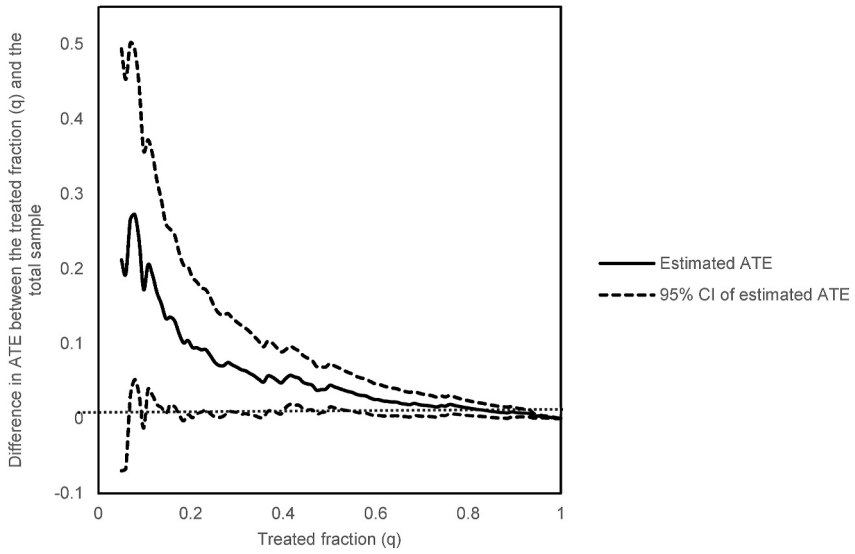


**FIGURE 1** Targeting Operator Characteristic curve for ATE in the test sample ($n$ = 1052). ATE, average treatment effect of exposure to high combat-related stressors (CRS) on persistent post-traumatic stress disorder (PTSD) defined as the difference in predicted probability of persistent PTSD in the presence versus absence of high CRS after adjusting for nonrandom exposure to high CRS with respect to the range of variables assessed in the baseline survey; Treated fraction (q), the proportion of soldiers who would either be prevented from being exposed to high CRS or protected from the effects of exposure by some to be determined intervention.

method for determining the optimal number of factors (Crawford et al., 2010) found that a one-factor model was optimal (factor loadings of 0.48–0.74). Based on this result, we standardized the frequency report for each stressor to a mean of 0 and variance of 1 and summed the 10 scores in the total sample (i.e., including soldiers in combat support and service support occupations as well as in combat arms) to create a summary scale. We then investigated the association between scores on this summary scale and prevalence of persistent T2-T3 PTSD in the total sample. The association was nonlinear, with by far the highest prevalence of persistent PTSD in the 30% of soldiers with highest summary scale scores. Based on this result, we defined high CRS as a dichotomy for being in the top three deciles of the total sample distribution. Prevalence of high CRS based on this definition was 40.8% (SE = 1.0) among combat arms soldiers.

## 3.3 | Estimated average effects of high CRS on persistent PTSD

The gross association of high CRS with persistent PTSD was estimated to be 7.3% (SE = 1.1%), $t = 6.9$, $p < 0.001$ prior to adjusting for nonrandom exposure to high CRS. The *grf* estimate of ATE, which adjusted for nonrandom exposure, was lower but still statistically significant: ATE = 6.7% (SE = 1.6%), $t = 4.3$, $p < 0.001$.

## 3.4 | Estimated heterogeneity in resilience to the effects of high CRS

Variation in ATE in the test sample based on predicted CATE (estimated in the training sample) was statistically significant, with AUTOC = 0.066 (SE = 0.031), $t = 2.1$, $p = 0.032$. Inspection of the TOC curve (Figure 1) shows that ATE was low for roughly 80% of the sample and dramatically higher for the 20% of least resilient soldiers. A more intuitive characterization of this difference can be had by using *grf* to estimate ATE separately in quintiles of the test sample based on predicted CATE scores (Table 3). ATE was estimated to be vanishingly small using this approach in the lowest quintile (ATE = 1.1%, SE = 1.6%, $t = 0.7$, $p = 0.48$), larger in the three intermediate quintiles (ATE = 3.2–5.6%, and ATE = 4.6%, SE = 1.6, $t = 2.9$, $p = 0.003$ pooled across the three quintiles), and by far largest in the highest quintile (ATE = 17.1%, SE = 5.5, $t = 3.1$, $p = 0.002$). Pooled ATE across the 80% of respondents excluding the least resilient (ATE = 3.8%, SE = 1.2, $t = 3.0$, $p = 0.003$) was only about half the size of ATE = 6.7% in the total sample, suggesting that prevalence of persistent PTSD might be cut in half if an intervention could be implemented to make the least resilient 20% of soldiers have resilience equal to the average in the remaining 80% of the sample. Pooled ATE in the total sample would be somewhat higher (ATE = 4.2%, SE = 1.0%, $t = 4.0$, $p = <0.001$) if an intervention could be implemented to make the least resilient 20% of soldiers have resilience equal to the average in the intermediate 60% of the sample.

### 3.4.1 | Predictor importance

The absolute SHAP value for the entire predictor set was 2.1%. This means that predicted CATE for a given individual would have changed by an average of 2.1% if all significant predictors were set to the sample mean rather than to their observed values. We define key predictors as those with proportional absolute SHAP values ($SHAP_P$)

**TABLE 3** Estimated average treatment effect (ATE) of high CRS on persistent PTSD in the total 40% test sample of enlisted PPDS soldiers with Combat Arms Military Occupational Specialties that completed both the T2 and T2 follow-up assessments ($n = 1052$) and in subsamples of the test sample defined by quintiles of predicted differential resilience based on a model estimated in the training sample.[a]

| | ATE % | (SE) % | *t* | *p* | (*n*) |
|---|---|---|---|---|---|
| Total test sample | 6.7 | (1.6) | 4.3 | <0.001 | (1052) |
| Resilience quintiles | | | | | |
| 1st (Most resilient) | 1.1 | (1.6) | 0.7 | 0.48 | (211) |
| 2nd | 5.1 | (2.2) | 2.3 | 0.024 | (210) |
| 3rd | 3.2 | (2.2) | 1.5 | 0.14 | (210) |
| 4th | 5.7 | (3.6) | 1.6 | 0.12 | (210) |
| 5th (Least resilient) | 17.1 | (5.5) | 3.1 | 0.002 | (211) |
| Other subsamples | | | | | |
| 1st–4th quintiles | 3.8 | (1.2) | 3.0 | 0.003 | (841) |
| 2nd–4th quintiles | 4.6 | (1.6) | 2.9 | 0.003 | (630) |
| 1st + (2nd—4th) × 4/3 | 4.2 | (1.0) | 4.0 | <0.001 | (841) |

Abbreviations: CRS, combat-related stressors; PPDS, pre-post deployment study; PTSD, post-traumatic stress disorder; SE, standard error of ATE.

[a]ATE was defined as the slope of persistent PTSD on high CRS in a regression model that adjusted for nonrandom exposure to high CRS with respect to measured pre-deployment predictors of persistent PTSD using the doubly robust method in the *grf R* package (Athey et al., 2024). The slope in the total test sample is 6.7%, which means that prevalence of persistent PTSD is 6.7% greater in the presence than absence of exposure to high CRS.

equal to at least 5% of this 2.1%. Ten such predictors were identified, but three of them had much larger $SHAP_P$ values (15.8%–47.2%) than the others (5.1%–6.7%) (Figure 2). The most important of the three was a composite measure of distress in the 30 days before deployment that combined standardized (to means of 0 and variances of 1) scores on separate scales of anxiety (Kessler, Calabrese, et al., 2013), depression (Kessler, Calabrese, et al., 2013), PTSD (Weathers et al., 1993), and symptoms of somatoform disorder (Spitzer et al., 1999). The $SHAP_P = 47.2\%$ of this composite predictor indicates that it explained close to half the variation in resilience accounted for by all baseline predictors. The bee swarm plot showed that the association of this predictor with CATE was largely positive; that is, that high symptom scores were associated with decreased resilience. It is noteworthy that the four symptom scales making up this composite measure were also included in the predictor set used to calculate SHAP values, but that only one of them, the anxiety scale ($SHAP_P = 6.7\%$), emerged as a key predictor. The dominant association of this predictor with CATE in the bee swarm plus was positive, indicating that pre-deployment anxiety was a more important component in decreased resilience than were the other aspects of distress.

The second most important predictor of CATE was the predicted probability of PTSD estimated in the training sample ($SHAP_P = 20.9\%$), again with a positive dominant direction of association with CATE (i.e., high predicted risk of PTSD was associated with decreased resilience). The third most important predictor was the soldier's number of lifetime internalizing disorders prior to deployment based on CIDI-SC scales (Kessler, Calabrese, et al., 2013) for DSM-IV diagnoses of major depressive disorder, bipolar spectrum disorder, including subthreshold BP disorder as defined elsewhere (Merikangas et al., 2007), generalized anxiety disorder, panic disorder (with or without agoraphobia), and PTSD ($SHAP_P = 15.8\%$). As with the other important predictors, the dominant direction of association with CATE in the bee swarm plot was positive (i.e., a high number of these disorders was associated with low resilience). It is noteworthy that none of the component lifetime disorders in this composite emerged as a key predictor. The other key predictors, each with a much smaller $SHAP_P$, included a pre-deployment report of the extent to which health problems interfered with the soldier's life, number of lifetime accidents-injuries, number of 12-month treatment visits for psychological problems, the personality characteristic of negative affectivity, number of firearms at home, and number of lifetime traumatic brain injuries. Each of these was associated with decreased resilience.

## 4 | DISCUSSION

We are aware of no previous study that created a composite resilience score using precision medicine analysis methods. We found using this approach that roughly 50% of the persistent PTSD among combat arms soldiers after returning from a tour of duty in Afghanistan occurred among the 20% of these soldiers judged by our model to be least resilient to the effects of high CRS. We found that these pre-deployment predictors of resilience were distinct from the pre-deployment predictors of persistent PTSD in that a composite measure of the latter sort accounted for only a minority ($SHAP_P = 20.6\%$) of the average absolute SHAP value of the resilience score. Instead, by far the most important predictor of differential resilience was a composite measure of pre-deployment distress in the month before deployment (SHAPP = 47.2%). It is noteworthy that we failed to find other significant predictors of differential resilience among those proposed as determinants of resilience in prior studies, such as childhood adversities (Carrasco-Barrios et al., 2020; Kraaijenvanger et al., 2020), but this could partly reflect the mediating effects of psychopathology on differential resilience given that most of the pre-deployment variables that had significant zero-order associations with predicted resilience, including several indicators of childhood adversity, were strongly correlated with pre-deployment measures of lifetime and recent psychopathology.

Our finding of significant variation in resilience is indirectly consistent with prior studies of PTSD that defined resilience as one of several dynamic symptom profiles derived from a latent growth mixture model in a sample of trauma-exposed respondents followed

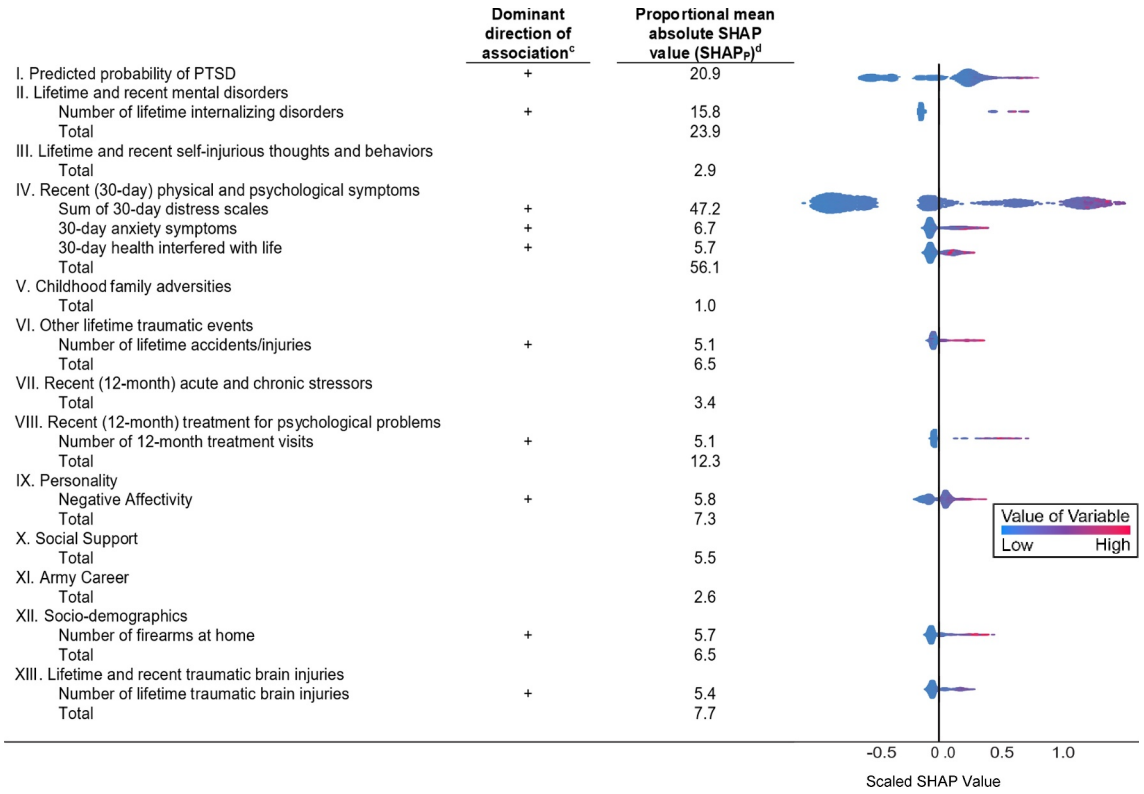


**FIGURE 2** Key predictors[a] (defined as those with SHAP$_P$ values of at least 5.0%) of differential resilience to the effects of high CRS on persistent post-deployment PTSD defined by SHAP value analysis among combat arms soldiers in the STARRS PPDS sample ($n = 2542$)[b]. [a]Key predictors are defined as those with mean absolute SHAP values at least 5% as large as the mean absolute SHAP value of the total model. The latter is 2.1%, which means that predicted CATE for a given individual would have changed by an average of 2.1% if all significant predictors were set to the sample mean rather than to the observed values. [b]The SHAP value analysis was carried out using the *xgboost* R package (Chen et al., 2024). The XGBoost algorithm (Chen & Guestrin, 2016) was used to predict individual differences in estimated CATE from the pre-deployment predictors that were used in grf (Athey et al., 2024). to generate individual-level estimates of CATE. Xgboost implements the tree SHAP method of estimating SHAP values (Lundberg & Lee, 2017). [c]The SHAP value of a predictor can vary across respondents whenever there are nonlinearities and/or interactions in the model. This variation can be characterized in a bee swarm plot in which the SHAP value for each respondent is treated as a dot and values of the SHAP value are defined by the *X* Axis. A red dot means a higher value of the predictor. A blue dot means a lower value. If the red dots are predominantly on the right side of the swarm, it means that high scores on the predictor are associated with higher CATEs, which is equivalent to lower resilience. [d]As noted above in FN a, a key predictor is defined as one with a mean absolute SHAP value at least 5% as large as the 2.1% mean absolute SHAP value of all predictors in the model. This means that any individual predictor with a mean absolute SHAP value of 0.01% (i.e., 5% of 2.1%) would be considered a key predictor. For ease of interpretation, we report proportional SHAP values (SHAP$_P$) directly in the table. For example, a predictor with a mean absolute SHAP value of, say, 0.01% would be reported as having a SHAP$_P$) of exactly 5.0% (i.e., 0.01/2.1). CRS, combat-related stressors; PTSD, post-traumatic stress disorder; SHAP, Shapley Additive Explanations; STARRS PPDS, Army Study to Assess Risk and Resilience in Servicemembers Pre-Post Deployment Study.

over time (LGMM; Kalisch et al., 2017). However, that alternative approach implicitly assumes that stress exposure either is constant across all subjects studied or can be controlled for additivity in a model designed to predict which trauma-exposed individuals will develop PTSD and will either recover, recover partially, or not recover at all. Our approach, in comparison, explicitly rejects the notion that an additive model is appropriate for studying resilience and attempts instead to identify stable non-additivities (i.e., interactions) between high CRS exposure and pre-deployment variables that modify the effects of high CRS in predicting persistent PTSD. The main conceptual difference between these two approaches is akin to the difference of randomization as between studying predictors of treatment response in an open trial (the LGMM approach) or a randomized controlled trial (our approach). In an open trial, symptomatic improvement can be caused by regression to the mean, spontaneous remission, or life experiences influenced by baseline predictors unrelated to the intervention (in our case, high CRS). In a controlled trial, in comparison, all these extraneous effects are controlled in expectation by randomization (or, in the case of observational studies, baseline covariate balancing), allowing both aggregate intervention effects and interactions of prescriptive predictors with the intervention to be seen more clearly.

Consistent with our innovative approach to defining resilience, our results are novel in finding powerful interactions between pre-deployment predictors and high CRS. This is intuitive from a stress diathesis perspective but leads to more nuance than in the LGMM approach and would consequently be expected to improve targeting

of interventions to prevent future PTSD (Schultebraucks et al., 2021). Furthermore, such interactions identify potential intervention strategies. Although we made no attempt to investigate the specific interactions making up the CRS, that type of more fine-grained analysis could be carried out using standard machine learning methods to examine the intervention implications of predictor importance profiles (Molnar, 2022).

We showed that multiple interactions can be aggregated into a stable composite to define resilience and create a parsimonious characterization of individual differences. For example, our results could be used to make a simple dichotomous distinction between the 20% of combat arms soldiers who would be non-resilient and those that would be resilient. In precision medicine, such composites are used to optimize treatment assignment. The same could be done here. For example, our resilience score could be used to target the 20% of non-resilient soldiers to special pre-deployment refresher sessions of Army Resilience Training (Hipps, 2017) or to match low-resilience soldiers with high-resilience soldiers as their battle buddies (Milzarski, 2021). Or, as unit commanders have latitude in deciding which soldiers to deploy and which to keep in the rear detachment, our predicted resilience score could be presented to unit commanders for them to use as they see fit in that decision-making process.

Our results are limited in four ways. First, as analysis was nonexperimental, unmeasured confounders could have introduced bias. Second, precision in estimating differential resilience was limited by the small number of soldiers in the sample with persistent PTSD and weak assessment of CRS. Future research working with a larger sample and more extensive combat stressor assessments (Kimbrel et al., 2014; Porter et al., 2018; Vogt et al., 2013) could take a data-driven approach to search for stressor clusters to improve the estimation of differential resilience. Third, by working with only three BCTs, we were unable to study unit-level characteristics predicting resilience (Campbell et al., 2019). Fourth, our resilience score was specific to the effects of high CRS on persistent PTSD. Different predictors might be involved in effects of other CRS (Monteith et al., 2018) on other post-redeployment outcomes of interest (Kwan et al., 2018; Regasa et al., 2016; Zuromski et al., 2020).

Within the context of these limitations, we found substantial variation in resilience to the effects of high CRS on persistent PTSD. We developed an optimized estimate of differential resilience with respect to the baseline predictors available to us that could be used to classify soldiers by resilience and target preventive interventions. Although predictors might differ for other stressor-outcome combinations (Southwick et al., 2014), and the predictors of differential intervention response could be different yet, our logic could be extended to consider these possibilities.

## AUTHOR CONTRIBUTIONS

**Ronald C. Kessler**: Conceptualization, Funding Acquisition, Investigation, Resources, Supervision, Validation, Writing—Original Draft, Writing—Review & Editing; **Robert M. Bossarte**: Conceptualization, Investigation, Project Administration, Resources, Writing—Review & Editing; **Irving Hwang**: Data Curation, Formal Analysis, Writing—Review & Editing; **Alex Luedtke**: Formal Analysis, Investigation, Methodology, Software, Supervision, Writing—Review & Editing; **James A. Naifeh**: Investigation, Writing—Review & Editing; **Matthew K. Nock**: Investigation, Writing—Review & Editing; **Maria Petukhova**: Data Curation, Formal Analysis, Validation, Writing—Review & Editing; **Ekaterina Sadikova**: Data Curation, Formal Analysis, Writing—Review & Editing; **Nancy A. Sampson**: Data Curation, Formal Analysis, Project Administration, Validation, Writing—Review & Editing; **Erik Sverdrup**: Formal Analysis, Software, Visualization, Writing—Review & Editing; **Jose R. Zubizarreta**: Investigation, Methodology, Software, Supervision, Visualization, Writing—Review & Editing; **Stefan Wager**: Investigation, Methodology, Project Administration, Resources, Software, Supervision, Writing—Review & Editing; **James Wagner**: Data Curation, Investigation, Project Administration, Writing—Review & Editing; **Murray B. Stein**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Resources, Writing—Review & Editing; **Robert J. Ursano**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Resources, Writing—Review & Editing.

## ACKNOWLEDGMENTS

The Army STARRS Team consists of Co-Principal Investigators: Robert J. Ursano, MD (Uniformed Services University) and Murray B. Stein, MD, MPH (University of California San Diego and VA San Diego Healthcare System). Site Principal Investigators: James Wagner, PhD (University of Michigan) and Ronald C. Kessler, PhD (Harvard Medical School). Army scientific consultant/liaison: Kenneth Cox, MD, MPH (Office of the Assistant Secretary of the Army (Manpower and Reserve Affairs). Other team members: Pablo A. Aliaga, MA (Uniformed Services University); David M. Benedek, MD (Uniformed Services University); Laura Campbell-Sills, PhD (University of California San Diego); Carol S. Fullerton, PhD (Uniformed Services University); Nancy Gebler, MA (University of Michigan); Meredith House, BA (University of Michigan); Paul E. Hurwitz, MPH (Uniformed Services University); Sonia Jain, PhD (University of California San Diego); Tzu-Cheg Kao, PhD (Uniformed Services University); Lisa Lewandowski-Romps, PhD (University of Michigan); Alex Luedtke, PhD (University of Washington and Fred Hutchinson Cancer Research Center); Holly Herberman Mash, PhD (Uniformed Services University); James A. Naifeh, PhD (Uniformed Services University); Matthew K. Nock, PhD (Harvard University); Nur Hani Zainal, PhD (Harvard Medical School); Nancy A. Sampson, BA (Harvard Medical School); and Alan M. Zaslavsky, PhD (Harvard Medical School). Army STARRS was sponsored by the Department of the Army and funded under cooperative agreement number U01MH087981 (2009–2015) with the National Institute of Mental Health (NIMH). Subsequently, STARRS-LS was sponsored and funded by the Department of Defense (USUHS grant number HU0001-15-2-0004). The contents are solely the responsibility of the authors and do not necessarily represent the views of the NIMH, the Department of the Army, the Department of Defense, or the Department of Veteran Affairs.

## CONFLICT OF INTEREST STATEMENT

In the past 3 years, Dr. Kessler was a consultant for Cambridge Health Alliance, Canandaigua VA Medical Center, Child Mind Institute, Holmusk, Massachusetts General Hospital, Partners Healthcare, Inc., RallyPoint Networks, Inc., Sage Therapeutics and University of North Carolina. He has stock options in Cerebral Inc., Mirah, Prepare Your Mind, Roga Sciences and Verisense Health. In the past 3 years, Dr. Stein received consulting income from Actelion, Acadia Pharmaceuticals, Aptinyx, atai Life Sciences, Boehringer Ingelheim, Bionomics, BioXcel Therapeutics, Clexio, EmpowerPharm, Engrail Therapeutics, GW Pharmaceuticals, Janssen, Jazz Pharmaceuticals, and Roche/Genentech. Dr. Stein has stock options in Oxeia Biopharmaceuticals and EpiVario. He is paid for his editorial work on *Depression and Anxiety* (Editor-in-Chief), *Biological Psychiatry* (Deputy Editor), and *UpToDate* (Co-Editor-in-Chief for Psychiatry). The remaining authors declare no competing interests.

## DATA AVAILABILITY STATEMENT

The STARRS Historical Administrative Data Study data used in this paper are not available for public release. DoD clearance is required to access this data. However, the Pre and Post Deployment data (PPDS) survey data are available to qualified researchers as restricted datasets through the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan (https://www.icpsr.umich.edu/web/ICPSR/studies/35197). Access to restricted datasets requires the applicant and their institution to establish confidential agreements with ICPSR. Researchers interested in this should submit applications via ICPSR's online Restricted Contracting System. Guidelines can be found under the data and documentation tab at the above URL.

## ETHICS STATEMENT

The study protocol was approved by the Research Ethics Committees of the Henry Jackson Foundation and Harvard Medical School (IRB15-0765) with a waiver of informed consent based on data being de-identified. Research has been performed in accordance with the Declaration of Helsinki.

## ORCID

*Ronald C. Kessler* https://orcid.org/0000-0003-4831-2305
*James A. Naifeh* https://orcid.org/0000-0001-7248-3054

## REFERENCES

Abadie, A., Chingos, M. M., & West, M. R. (2018). Endogenous stratification in randomized experiments. *The Review of Economics and Statistics*, *100*(4), 567–580. https://doi.org/10.1162/rest_a_00732

Athey, S., Friedberg, R., Hadad, V., Hirshberg, D., Miner, L., Sverdrup, E., Wager, S., & Wright, M. (2024). Generalized random forests: Package 'grf'. Retrieved September 1, 2024 from https://cran.r-project.org/web/packages/grf/grf.pdf

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, *47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Benjet, C., Bromet, E., Karam, E. G., Kessler, R. C., McLaughlin, K. A., Ruscio, A. M., Shahly, V., Stein, D. J., Petukhova, M., Hill, E., Alonso, J., Atwoli, L., Bunting, B., Bruffaerts, R., Caldas-de-Almeida, J. M., de Girolamo, G., Florescu, S., Gureje, O., Huang, Y., …, & Koenen, K. C. (2016). The epidemiology of traumatic event exposure worldwide: Results from the World mental health survey Consortium. *Psychological Medicine*, *46*(2), 327–343. https://doi.org/10.1017/s0033291715001981

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders-fifth edition (PCL-5) in veterans. *Psychological Assessment*, *28*(11), 1379–1391. https://doi.org/10.1037/pas0000254

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Campbell, J. S., Wallace, M. L., Germain, A., & Koffman, R. L. (2019). A predictive analytic approach to planning combat stress control operations. *J Stress Manag*, *26*(2), 120–131. https://doi.org/10.1037/str0000092

Carrasco-Barrios, M. T., Huertas, P., Martín, P., Martín, C., Castillejos, M. C., Petkari, E., & Moreno-Küstner, B. (2020). Determinants of suicidality in the European general population: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, *17*(11), 4115. https://doi.org/10.3390/ijerph17114115

Chattopadhyay, A., Hase, C. H., & Zubizarreta, J. R. (2020). Balancing vs modeling approaches to weighting in practice. *Statistics in Medicine*, *39*(24), 3227–3254. https://doi.org/10.1002/sim.8659

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *11*, 785–794. https://dl.acm.org/doi/10.1145/2939672.2939785

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Che, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., & Yuan, J. (2024). xgboost: Extreme gradient boosting. In (version 1.7.8.1). https://github.com/dmlc/xgboost

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, *70*(6), 885–901. https://doi.org/10.1177/0013164410379332

Desai, R. J., & Franklin, J. M. (2019). Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: A primer for practitioners. *BMJ*, *367*, l5657. https://doi.org/10.1136/bmj.l5657

Doody, C. B., Robertson, L., Cox, K. M., Bogue, J., Egan, J., & Sarma, K. M. (2021). Pre-deployment programmes for building resilience in military and frontline emergency service personnel. *Cochrane Database of Systematic Reviews*, *12*(12), Cd013242. https://doi.org/10.1002/14651858.CD013242.pub2

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, *173*(7), 761–767. https://doi.org/10.1093/aje/kwq439

Hernán, M. A., & Robins, J. M. Causal inference: What if (1 ed.). CRC Press. (In Press).

Hines, L. A., Sundin, J., Rona, R. J., Wessely, S., & Fear, N. T. (2014). Posttraumatic stress disorder post Iraq and Afghanistan: Prevalence among military subgroups. *Canadian Journal of Psychiatry*, *59*(9), 468–479. https://doi.org/10.1177/070674371405900903

Hipps, T. (2017). *CSF2 training points Soldiers toward readiness and resilience.* U.S. Army. Retrieved from https://www.army.mil/article/181247/csf2_training_points_soldiers_toward_readiness_and_resilience

Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, *351*(1), 13–22. https://doi.org/10.1056/NEJMoa040603

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical Sciences: An introduction.* Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Jones, E. (2006). Historical approaches to post-combat disorders. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, *361*(1468), 533–542. https://doi.org/10.1098/rstb.2006.1814

Kalisch, R., Baker, D. G., Basten, U., Boks, M. P., Bonanno, G. A., Brummelman, E., Chmitorz, A., Fernàndez, G., Fiebach, C. J., Galatzer-Levy, I., Geuze, E., Groppa, S., Helmreich, I., Hendler, T., Hermans, E. J., Jovanovic, T., Kubiak, T., Lieb, K., Lutz, B., …, & Kleim, B. (2017). The resilience framework as a strategy to combat stress-related disorders. *Nat Hum Behav*, *1*(11), 784–790. https://doi.org/10.1038/s41562-017-0200-8

Kennedy, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic J Stat*, *17*(2), 3008–3049. https://doi.org/10.1214/23-EJS2157

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Benjet, C., Bromet, E. J., Cardoso, G., Degenhardt, L., de Girolamo, G., Dinolova, R. V., Ferry, F., Florescu, S., Gureje, O., Haro, J. M., Huang, Y., Karam, E. G., Kawakami, N., Lee, S., Lepine, J. P., Levinson, D., …, & Koenen, K. C. (2017). Trauma and PTSD in the WHO World mental health surveys. *European Journal of Psychotraumatology*, *8*(sup5), 1353383. https://doi.org/10.1080/20008198.2017.1353383

Kessler, R. C., Calabrese, J. R., Farley, P. A., Gruber, M. J., Jewell, M. A., Katon, W., Keck, P. E., Nierenberg, A. A., Sampson, N. A., Shear, M. K., Shillington, A. C., Stein, M. B., Thase, M. E., & Wittchen, H. U. (2013a). Composite International Diagnostic Interview screening scales for DSM-IV anxiety and mood disorders. *Psychological Medicine*, *43*(8), 1625–1637. https://doi.org/10.1017/s0033291712002334

Kessler, R. C., Colpe, L. J., Fullerton, C. S., Gebler, N., Naifeh, J. A., Nock, M. K., Sampson, N. A., Schoenbaum, M., Zaslavsky, A. M., Stein, M. B., Ursano, R. J., & Heeringa, S. G. (2013b). Design of the Army study to assess risk and resilience in servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research*, *22*(4), 267–275. https://doi.org/10.1002/mpr.1401

Kessler, R. C., Heeringa, S. G., Colpe, L. J., Fullerton, C. S., Gebler, N., Hwang, I., Naifeh, J. A., Nock, M. K., Sampson, N. A., Schoenbaum, M., Zaslavsky, A. M., Stein, M. B., & Ursano, R. J. (2013c). Response bias, weighting adjustments, and design effects in the Army study to assess risk and resilience in servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research*, *22*(4), 288–302. https://doi.org/10.1002/mpr.1399

Kessler, R. C., Santiago, P. N., Colpe, L. J., Dempsey, C. L., First, M. B., Heeringa, S. G., Stein, M. B., Fullerton, C. S., Gruber, M. J., Naifeh, J. A., Nock, M. K., Sampson, N. A., Schoenbaum, M., Zaslavsky, A. M., & Ursano, R. J. (2013d). Clinical reappraisal of the composite international diagnostic interview screening scales (CIDI-SC) in the Army study to assess risk and resilience in servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research*, *22*(4), 303–321. https://doi.org/10.1002/mpr.1398

Kimbrel, N. A., Evans, L. D., Patel, A. B., Wilson, L. C., Meyer, E. C., Gulliver, S. B., & Morissette, S. B. (2014). The critical warzone experiences (CWE) scale: Initial psychometric properties and association with PTSD, anxiety, and depression. *Psychiatry Research*, *220*(3), 1118–1124. https://doi.org/10.1016/j.psychres.2014.08.053

Kok, B. C., Wilk, J. E., Wickham, R. E., Bongar, B., Riviere, L. A., & Brown, L. M. (2020). Military occupation as a moderator between combat exposure and posttraumatic stress disorder symptoms in US Army personnel. *Military Psychology*, *32*(5), 410–418. https://doi.org/10.1080/08995605.2020.1782625

Kraaijenvanger, E. J., Pollok, T. M., Monninger, M., Kaiser, A., Brandeis, D., Banaschewski, T., & Holz, N. E. (2020). Impact of early life adversities on human brain functioning: A coordinate-based meta-analysis. *Neuroscience & Biobehavioral Reviews*, *113*, 62–76. https://doi.org/10.1016/j.neubiorev.2020.03.008

Kwan, J., Jones, M., Somaini, G., Hull, L., Wessely, S., Fear, N. T., & MacManus, D. (2018). Post-deployment family violence among UK military personnel. *Psychological Medicine*, *48*(13), 2202–2212. https://doi.org/10.1017/s0033291717003695

Leppin, A. L., Bora, P. R., Tilburt, J. C., Gionfriddo, M. R., Zeballos-Palacios, C., Dulohery, M. M., Sood, A., Erwin, P. J., Brito, J. P., Boehmer, K. R., & Montori, V. M. (2014). The efficacy of resiliency training programs: A systematic review and meta-analysis of randomized trials. *PLoS One*, *9*(10), e111420. https://doi.org/10.1371/journal.pone.0111420

Luedtke, A. R., & van der Laan, M. J. (2017). Evaluating the impact of treating the optimal subgroup. *Statistical Methods in Medical Research*, *26*(4), 1630–1640. https://doi.org/10.1177/0962280217708664

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. Curran Associates Inc. Retrieved from https://dl.acm.org/doi/10.5555/3295222.3295230

Maercker, A., Cloitre, M., Bachem, R., Schlumpf, Y. R., Khoury, B., Hitchcock, C., & Bohus, M. (2022). Complex post-traumatic stress disorder. *Lancet*, *400*(10345), 60–72. https://doi.org/10.1016/s0140-6736(22)00821-2

Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M., Petukhova, M., & Kessler, R. C. (2007). Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Archives of General Psychiatry*, *64*(5), 543–552. https://doi.org/10.1001/archpsyc.64.5.543

Milzarski, E. (2021). 5 reasons why the battle buddy system was secretly brilliant. We Are The Mighty. Retrieved September 1, 2024 from https://www.wearethemighty.com/mighty-culture/merits-of-battle-buddy-system/?rebelltitem=4#rebelltitem4

Molnar, C. (2022). Interpretable machine learning: A guide for making Black box models explainable (2nd ed.). Retrieved from https://christophm.github.io/interpretable-ml-book/

Monteith, L. L., Hoffmire, C. A., Holliday, R., Park, C. L., Mazure, C. M., & Hoff, R. A. (2018). Do unit and post-deployment social support influence the association between deployment sexual trauma and suicidal ideation? *Psychiatry Research*, *270*, 673–681. https://doi.org/10.1016/j.psychres.2018.10.055

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *J R Stat Soc, Ser B Methodol*, *65*(2), 331–355. https://doi.org/10.1111/1467-9868.00389

Papini, S., Norman, S. B., Campbell-Sills, L., Sun, X., He, F., Kessler, R. C., Ursano, R. J., Jain, S., & Stein, M. B. (2023). Development and validation of a machine learning prediction model of posttraumatic stress disorder after military deployment. *JAMA Network Open*, *6*(6), e2321273. https://doi.org/10.1001/jamanetworkopen.2023.21273

Paulson, D. S., & Krippner, S. (2007). *Haunted by combat: Understanding PTSD in war veterans including women, reservists, and those coming back from Iraq*. Bloomsbury Publishing. Retrieved from https://books.google.com/books?id=gJ2Mq7nEZ8MC

Porter, B., Hoge, C. W., Tobin, L. E., Donoho, C. J., Castro, C. A., Luxton, D. D., & Faix, D. (2018). Measuring aggregated and specific combat exposures: Associations between combat exposure measures and posttraumatic stress disorder, depression, and alcohol-related problems. *Journal of Traumatic Stress*, *31*(2), 296–306. https://doi.org/10.1002/jts.22273

R Core Team. (2018). *R: A language and environment for statistical computing*. Foundation for Statistical Computing. Retrieved April 28, 2023 from https://www.r-project.org

Regasa, L. E., Thomas, D. M., Gill, R. S., Marion, D. W., & Ivins, B. J. (2016). Military deployment may increase the risk for traumatic brain injury following deployment. *The Journal of Head Trauma Rehabilitation*, *31*(1), E28–E35. https://doi.org/10.1097/htr.0000000000000155

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In D. Y. Lin & P. J. Heagerty (Eds.), *Proceedings of the second Seattle symposium in biostatistics: Analysis of correlated data* (pp. 189–326). Springer. https://doi.org/10.1007/978-1-4419-9076-1_11

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Sabé, M., Chen, C., El-Hage, W., Leroy, A., Vaiva, G., Monari, S., Premand, N., Bartolomei, J., Caiolo, S., Maercker, A., Pietrzak, R. H., Cloître, M., Kaiser, S., & Solmi, M. (2024). Half a century of research on post-traumatic stress disorder: A scientometric analysis. *Current Neuropharmacology*, *22*(4), 736–748. https://doi.org/10.2174/1570159x22666230927143106

SAS Software 9.4 edn. (2013). SAS Insititue Inc.

Schultebraucks, K., Qian, M., Abu-Amara, D., Dean, K., Laska, E., Siegel, C., Gautam, A., Guffanti, G., Hammamieh, R., Misganaw, B., Mellon, S. H., Wolkowitz, O. M., Blessing, E. M., Etkin, A., Ressler, K. J., Doyle, F. J., 3rd, Jett, M., & Marmar, C. R. (2021). Pre-deployment risk factors for PTSD in active-duty personnel deployed to Afghanistan: A machine-learning approach for analyzing multivariate predictors. *Molecular Psychiatry*, *26*(9), 5011–5022. https://doi.org/10.1038/s41380-020-0789-2

Southwick, S. M., Bonanno, G. A., Masten, A. S., Panter-Brick, C., & Yehuda, R. (2014). Resilience definitions, theory, and challenges: Interdisciplinary perspectives. *European Journal of Psychotraumatology*, *5*(1). https://doi.org/10.3402/ejpt.v5.25338

Spitzer, R. L., Kroenke, K., & Williams, J. B., & the Patient Health Questionnaire Primary Care Study Group. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA*, *282*(18), 1737–1744. https://doi.org/10.1001/jama.282.18.1737

Sverdrup, E., Petukhova, M., & Wager, S. (In Press).Estimating treatment effect heterogeneity in Psychiatry: A review and tutorial with causal forests. *International Journal of Methods in Psychiatric Research*.

Ursano, R. J., Colpe, L. J., Heeringa, S. G., Kessler, R. C., Schoenbaum, M., & Stein, M. B. (2014). The Army study to assess risk and resilience in servicemembers (Army STARRS). *Psychiatry*, *77*(2), 107–119. https://doi.org/10.1521/psyc.2014.77.2.107

U.S. Department of Veterans Affairs. (2023). Corporate data warehouse (CDW). Retrieved February 1 2023 from https://www.hsrd.research.va.gov/for_researchers/vinci/cdw.cfm

VanderWeele, T. J., Luedtke, A. R., van der Laan, M. J., & Kessler, R. C. (2019). Selecting optimal subgroups for treatment using many covariates. *Epidemiology*, *30*(3), 334–341. https://doi.org/10.1097/ede.0000000000000991

Varadhan, R., & Seeger, J. (2013). Estimation and reporting of heterogeneity of treatment effects. In P. Velentgas, N. Dreyer, P. Nourjah, S. Smith, & M. Torchia (Eds.), *Developing a protocol for observational comparative effectiveness research: A user's guide*. Agency for Healthcare Research and Quality. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK126188/

Vogt, D., Smith, B. N., King, L. A., King, D. W., Knight, J., & Vasterling, J. J. (2013). Deployment risk and resilience inventory-2 (DRRI-2): An updated tool for assessing psychosocial risk and resilience factors among service members and veterans. *Journal of Traumatic Stress*, *26*(6), 710–717. https://doi.org/10.1002/jts.21868

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Weathers, F., Litz, B., Herman, D., Huska, J. A., & Keane, T. (1993). The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. In *Annual convention of the international society for traumatic stress studies*. Retrieved from https://www.researchgate.net/publication/291448760_The_PTSD_Checklist_PCL_Reliability_validity_and_diagnostic_utility

Wild, J., Smith, K. V., Thompson, E., Béar, F., Lommen, M. J., & Ehlers, A. (2016). A prospective study of pre-trauma risk factors for post-traumatic stress disorder and depression. *Psychological Medicine*, *46*(12), 2571–2582. https://doi.org/10.1017/s0033291716000532

Xue, C., Ge, Y., Tang, B., Liu, Y., Kang, P., Wang, M., & Zhang, L. (2015). A meta-analysis of risk factors for combat-related PTSD among military personnel and veterans. *PLoS One*, *10*(3), e0120270. https://doi.org/10.1371/journal.pone.0120270

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., & Wager, S. (2024). Evaluating treatment prioritization rules via rank-weighted average treatment effects. *J Am Stat*, 1–25. https://doi.org/10.1080/01621459.2024.2393466

Zuromski, K. L., Bernecker, S. L., Chu, C., Wilks, C. R., Gutierrez, P. M., Joiner, T. E., Liu, H., Naifeh, J. A., Nock, M. K., Sampson, N. A., Zaslavsky, A. M., Stein, M. B., Ursano, R. J., Kessler, R. C., Steven Heeringa, Wagner, J., Cox, K., & Aliaga, P. A. (2020). Pre-deployment predictors of suicide attempt during and after combat deployment: Results from the Army Study to Assess Risk and Resilience in Servicemembers. *J Psychiatr Res*, *121*, 214–221. https://doi.org/10.1016/j.jpsychires.2019.12.003

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kessler, R. C., Bossarte, R. M., Hwang, I., Luedtke, A., Naifeh, J. A., Nock, M. K., Petukhova, M., Sadikova, E., Sampson, N. A., Sverdrup, E., Zubizarreta, J. R., Wager, S., Wagner, J., Stein, M. B., & Ursano, R. J. (2024). A prediction model for differential resilience to the effects of combat-related stressors in US army soldiers. *International Journal of Methods in Psychiatric Research*, e70006. https://doi.org/10.1002/mpr.70006